

POL 213: Exam 1

Instructions: This exam will focus on implementation of the logit and probit models for binary response variables and on MLE. This exam is due Monday, June 1 by 9:00 AM. (although you will have a problem set assigned in interim, so do not wait until last moment). All analysis should be done in R. The complete R output should be appended to the exam and appropriately annotated. For purposes of this exam, you will need to do the following:

1. Find a political science data set with a binary response variable and four covariates, two of which should be non-binary. The response variable should be of reasonable interest to political scientists (i.e., do not model the probability of being a man or woman [though you might model the probability of a male vs. female incumbent]). The covariates should be theoretically connected to the response variable in a non-haphazard way. The sample size should be sufficiently large to proceed with MLE. Once you have compiled the data, please describe the data as you would for a substantive paper. Explain the hypotheses with respect to the covariates, and then estimate both a logit and probit model.

For both models, please do the following:

- a. Present the results in sensible and readable tabular form using appropriate labeling. The table should be suitable for a normal research paper.
- b. Compute $\Pr(Y = 1)$ for three covariate profiles along with the standard errors, giving a sentence or two explanation as to what each probability profile means.

For the logit model, please do the following:

- c. Give a full interpretation of the results of the logit model. What is the degree of uncertainty around the parameter estimates. In interpreting the model, give at least one table of interpretable results that is sensible and easy to read. Additionally, provide at least one figure that visually describes the relationship between one (or more) of the covariates and some quantity of interest (like $\Pr(Y = 1)$ or marginal change in the probability. Along with this, discuss the uncertainty around these quantities. The interpretation should be written as if this were the “analysis” section of a research paper. As such, it will take several pages to do this appropriately (perhaps 4-6 pages). In the interpretation, I will need to see illustration of the following:
 - i. marginal effects or first difference for at least one continuous covariate (along with standard errors).
 - ii. sensible odds ratio interpretation for at least one binary and one continuous covariate (along with standard errors).

Presentation and interpretation of results are fundamentally important. This first question will count for 70 percent of the exam grade.

2. Please precisely explain why the OLS model is problematic with respect to binary response variables. Explain next why logit and probit functions overcome these problems. Give as much mathematic detail as you can. Please be careful in writing out mathematical expressions. Make sure you get the notation correct. This question will account for 10 percent of the exam grade.

3. Using the data from question 1, write out the log-likelihood function for a probit model and optimize the function. Verify your results are identical to R's `glm`. For this problem, you may use just a single covariate. For this question, you may work with other students on R coding. This holds only for this question. This question will account for 10 percent of the exam grade.

4. Please answer the following questions: a. Under what conditions might the first order condition hold, but a true maximum is not obtained? What are possible solutions?

b. Under what conditions might the second order condition not hold? What are some possible solutions?

c. For what reasons might a likelihood function take many iterations to converge?

d. If you had to do everything on your own, how would you know your model converged?

This question is worth 10 percent of the exam.